Method and Apparatus for Synchronizing Audio and Video in Encrypted Videoconferences

BACKGROUND OF THE INVENTION

TECHNICAL FIELD

The invention relates to videoconferencing systems. More particularly, the invention relates to a method and apparatus for synchronizing audio and video in encrypted videoconferences.

DESCRIPTION OF THE PRIOR ART

In many video conferencing systems, it is possible to conduct a conference involving more than two conference sites. In such conferences, the network topology often incorporates a hub that receives incoming audio and video signals from each of the participating sites, and routes appropriate outgoing audio and video signals to each site. Because each site typically has a single display on which to present video signals routed from the hub, a single video signal is routed from the hub to each site to conserve bandwidth. However, unlike video, audio for more than one site may be presented simultaneously at a given site, and indeed conference participants at a

given site viewing a single video signal may still benefit from hearing audio originating from all conference sites.

Existing systems meet this need by mixing audio signals and selecting video signals at the hub. All audio signals received at the hub are mixed together and routed to each site. However, only the video signal that a particular site is to display is routed to that particular site. The audio mixing and video selection operations are sufficiently simple that the latencies introduced into the audio and video signals are comparable. The audio and video presented at the destination site are therefore synchronized.

In the case of a video conferencing system incorporating encryption, several challenges are encountered. If the standard approach is to be used, the video and audio signals must be decrypted and decompressed prior to audio mixing and video selection. This leads to a substantial increase in latency. Further, it requires that the physical site housing the hub be secured and authorized to handle unencrypted information.

An alternative approach involves sending the audio signal received from each site to each other site. However, in this approach each site must then decrypt and decompress the audio and video signals separately. Most notably, the audio signal originating from the same site as the displayed video is handled separately from the displayed video. The discrepancy in latencies that results produces a desynchronization of the audio associated with the displayed video. The result is a confusing, distracting, and unsatisfying experience for the conference participants.

It would be advantageous to provide a system that preserves the synchronization of the audio and video presented at a secure conferencing site without necessitating decryption, decompression, compression, and encryption of signals at the hub.

SUMMARY OF THE INVENTION

The invention provides a system that preserves the synchronization of the audio and video presented at a secure conferencing site without necessitating decryption, decompression, compression, and encryption of signals at the hub. The presently preferred embodiment of the invention provides an apparatus and method for synchronizing audio and video in encrypted videoconferences that comprises a plurality of conference sites; and a hub for receiving a compressed and encrypted composite audio and video signal from each site, determining for each conference site a currently displayed composite audio and video signal, and transmitting each currently displayed composite audio and video signal to each respective site; said hub receiving a compressed and encrypted audio only signal from each site; wherein said hub routes all incoming compressed and encrypted audio only signals to each site. The invention further comprises an audio deselection and mixing device located at each conference site that deselects the audio only signal corresponding to the currently displayed composite audio and video signal and mixes all other audio only signals with the audio signal within the currently displayed composite audio and video signal.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block schematic diagram showing a system that implements a method and apparatus for synchronizing audio and video in encrypted videoconferences according to the invention; and

Fig. 2 is a block schematic diagram showing a video conference location that operates in connection with a method and apparatus for synchronizing audio and video in encrypted videoconferences according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides a system that preserves the synchronization of the audio and video presented at a secure conferencing site without necessitating decryption, decompression, compression, and encryption of signals at the hub.

Fig. 1 is a block schematic diagram showing a system that implements a method and apparatus for synchronizing audio and video in encrypted videoconferences according to the invention. In the preferred embodiment of the herein disclosed conferencing system, each of sites A-E, 11, 13, 15, 17, and 19, respectively, sends to the hub 10 a compressed and encrypted, composite audio and video signal. For each of the sites, the hub determines a currently displayed composite audio and video signal, based upon conference control information, and sends this composite audio and video signal to each respective site without decompressing or decrypting

the signal. There is no global active site. Instead, it is unique to each site. Thus, each site gets its own currently displayed composite signal.

Each site also sends to the hub a compressed and encrypted audio only signal. It should be noted that the audio only signal sent from each site may in fact be a mixed audio signal composed of audio obtained from several microphones at a single conferencing site. The hub routes all of the incoming compressed and encrypted audio only signals to each site.

Fig. 2 is a block schematic diagram showing a video conference location that operates in connection with a method and apparatus for synchronizing audio and video in encrypted videoconferences according to the invention. Each site, such as the five seat audio-video teleconference center 11 shown in Fig. 2a, decrypts, decompresses, and then displays the video within the composite audio and video signal received from the hub. The actual technique used for encryption/decryption and compression/decompression is a matter of choice to the person skilled in the art and is, therefore, not discussed in detail herein.

The signals transmitted to and from each site typically comprise conference control signals 22 to coordinate feeds and switching via an out-of-band mechanism such as an intranet or the Internet; a locally selected compressed and encrypted composite audio and video output 23; a compressed and encrypted audio only output preferably obtained by mixing several microphone feeds obtained at the site 24; a compressed and encrypted primary view composite audio and video input 25 selected by the hub control; a compressed and encrypted secondary view composite audio and video input 26 selected by the hub control for split screen generation (see the discussion

below); and n lines of compressed and encrypted audio only inputs 27 which correspond to each site in the conference.

The audio from the composite audio and video input signal, together with the other, separately decrypted and decompressed audio only input signals, is passed to an audio deselection and mixing device 21 (Fig. 2b). The separate audio only signal corresponding to the audio signal within the composite audio and video input signal is deselected by the device using a logic control signal 28 generated by an executive controller 12 (see Fig. 1). The logic control signal is shown in Fig. 1 as an out-of-band signal C2 generated by the executive controller, *i.e.* the hub controller, based upon video selection signals within the system. See Table 1 below, which details this exemplary audio selection logic scheme. Note that Table 1 shows the audio from the composite audio and video signal for the sending room in an upper cell of each receiving room row and the combined audio only signals from which the sending room audio has been subtracted in a lower cell of each receiving room row. For example, the rows for receiving room A intersects a column for sending room B in which the audio from the composite audio and video signals for sending room B is shown in an upper cell and the combined audio only signal from which the audio for room B has been subtracted, i.e. rooms CDE, shown in a lower cell. Those skilled in the art will appreciate that any known technique may be used for the audio deselection process.

The other audio signals, including the audio from within the composite signal, are mixed together and reproduced at the conferencing site. This process ensures that each audio signal is reproduced only once. Because the audio and video within the

composite audio and video signal are transmitted, decrypted, and decompressed together, the latencies introduced into the signals are well matched.

Table 1. Audio Selection Logic

RECEIVING ROOM	SENDING ROOM A	SENDING ROOM B	SENDING ROOM C	SENDING ROOM D	SENDING ROOM E
A	-	B	C	D	E
	-	CDE	BDE	BCE	BCD
B	A	-	C	D	E
	CDE	-	ADE	ACE	ACD
C	A	B	-	D	E
	BDE	ADE	-	ABE	ABD
D	A	B	C	-	E
	BCE	ACE	ABE	-	ABC
E	A	B	C	D	-
	BCD	ACD	ABD	ABC	-

The audio associated with the displayed video is therefore synchronized with the displayed video. Because the audio signals transmitted separately are processed separately, a latency different from that of the composite signal may be introduced. However, because these audio signals are not associated with the video displayed, this discrepancy is not noticeable to the participants. Nonetheless, the audio

deselection device may be equipped with delay circuitry to attempt to better align the separate audio signals with the composite signal.

If a split screen display is to be presented at a site, the hub transmits two composite audio and video signals to the site. Following decryption and decompression of the composite signals, the site uses a split screen composition processor to compose the split screen display from two video signals. In this case, two audio signals are deselected using the audio deselection device 21.

The audio deselection hub may also be used to deselect those audio signals not directly associated with the ongoing conversation. This may help in reducing the sense of background noise and audio clutter often observed during conferences where several audio signals are mixed.

Although the invention is described herein with reference to the preferred embodiment, one skilled in the art will readily appreciate that other applications may be substituted for those set forth herein without departing from the spirit and scope of the invention.

Notably, while the invention is describe with respect to a secure conferencing system incorporating both compression and encryption, the invention is also useful in conferencing systems incorporating only encryption, only compression, and neither encryption nor compression. In systems incorporating only encryption, the invention obviates the need for securing the conference hub. In systems incorporating only compression, the invention reduces the total system latency. In systems

incorporating neither encryption nor compression, the invention ensures optimal synchronization of audio and video signals.

Accordingly, the invention should only be limited by the Claims included below.